DOCUMENT RESUME

ED 277 239 FL 016 168

AUTHOR Loman, Bengt
TITLE Reports on Swedish Projects Relevant to the Study of Impromptu Speech: Talsyntax.
PUB DATE 82
NOTE 9p.; In: Enkvist, Nils Erik, Ed. Impromptu Speech: A Symposium. Papers Contributed to a Symposium on Problems in the Linguistic Study of Impromptu Speech (Abo, Finland, November 20-22, 1981); see FL 016 148.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classification; Data Collection; Discourse Analysis; Foreign Countries; Language Research; *Language Variation; Native Speakers; *Oral Language; Program Administration; *Program Design; *Research Methodology; Research Projects; Speech Acts; Spontaneous Behavior; Standard Spoken Usage; Structural Analysis (Linguistics); *Swedish; Syntax; Uncommonly Taught Languages
IDENTIFIERS *Spontaneous Speech; Sweden

ABSTRACT

The project Talsyntax (Syntax of Spoken Swedish), part of a large-scale study of modern Swedish language usage, examined the general characteristics and peculiarities of spoken Swedish. Three important aspects of the study are data collection, methodological results, and empirical results. The data collection effort has produced an archive of representative samples of spoken Swedish, with special attention given to social, regional, and situational variations. Most of the tape-recorded material has been transcribed, grammatically analyzed, and made available to researchers. Important methodological aspects of the study involved computer processing of the data collected, segmentation, and the development of methods for collecting and classifying the material according to nonlinguistic categories such as informant type and situation. The study's empirical results are many, but a general observation is that spoken Swedish is surprisingly homogeneous in the areas studied and more rule-governed than anticipated. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED277239

REPORTS ON SWEDISH PROJECTS RELEVANT TO THE STUDY OF IMPROMPTU SPEECH

I. Talsyntax

Bengt Loman
Abo Akademi

FL 016 168

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

■ This document has been reproduced as received from the person or organization originating it.

□ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Enkvist

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

REPORTS ON SWEDISH PROJECTS RELEVANT TO THE STUDY OF IMPROMPTU SPEECH

I. Talsyntax

Bengt Loman
Åbo Akademi

The project *Talsyntax* (Syntax of Spoken Swedish) was started in 1966 at the University of Lund as a branch of a large-scale research program called *Svenska språket i nutidens samhälle* (Swedish in today's society). The program comprised several Swedish universities. The different projects within the program had a common goal in describing the use of modern Swedish on the basis of large materials of data. The issue was to investigate how people really write and speak, and present the results as much as possible in a statistical format. This goal and methodology demanded a research organization on a broad basis, and with a long-range perspective. Grants from organizations like *Riksbankens jubileumsfond* – which took care of the financial aspects of *Talsyntax* in their entirety – were to provide the means for fulfilling the ambitious goal.

The project was set up at a time when the grammar of Standard Swedish and of the Swedish dialects had long been considered respectable fields of research. However, there had been few if any attempts to describe spontaneous speech and its ranges of individual, social and situational variation. As regards syntax, it was also desirable to enrich our knowledge of the relation between the norm, as manifested in the school-grammars for instance, and the everyday use of language. A number of similar problems that had not previously been dealt with constituted the point of departure for our work.

In the following I will try to summarize the results. The three main aspects that will be dealt with are: (1) data collecting, (2) methodological results and (3) empirical results.

1. Data Collecting

One of the main results of the work is the establishment of an archive of representative samples of spoken Swedish. The samples are based on the standard language, but special attention has been paid to social, regional and situational variation. Basically, the material consists of tape recordings, most of which (about one million words) have been transcribed. Moreover, about 75% of the material has undergone a very narrow transcription, copies of which are available to researchers. More important, however, is the fact that the greater part of the material has been grammatically analyzed. The result of this analysis, together with the texts themselves, is available on magnetic tape for computer processing. These magnetic tapes contain detailed information on the occurrence and frequency of different syntactic phenomena in different types of texts, also with respect to different types of speakers. It is furthermore possible to extract large numbers of examples for more detailed investigation, these examples being available through a mechanical print-out system. So far about 125.000 words from the material have been transferred onto magnetic tape. A corpus of written Swedish of some 235.000 words is also available on magnetic tape, as during part of its existence the *Talsyntax* project has been coordinated with the project *Skrivsyntax* (Syntax of Written Swedish). The latter has been financed by the Swedish School Board, *Skolöverstyrelsen*, and its main goal has been to analyze and evaluate the written language of high-school students on the basis of professionally written prose. Together the spoken and written material constitutes a corpus that is unique in the world: what we have collected on magnetic tape is not only a large amount of texts, but also detailed grammatical analyses of each sentence in the different texts. It is desirable that this corpus should be maintained as a living archive, which can give researchers rich and easily available data. For this purpose the Department of Nordic Languages at the University of Lund has formed a follow-up project called *Talbanken* (The Speech Bank).

2. Methodological Results

The ultimate goal of the research was to make the linguistic description as complete as possible. The aim was not to investigate specific details in the use of Swedish, such as the use of the genitive form or the passive construction. The aim was rather to arrive at a description that went from the higher-level syntactic units, the macrosyntagms, down to the specific parts of speech and their components. The description was also to take into account different prosodic phenomena (stress and intonation) and aspects of speech production itself (including false starts and repetitions, which are so characteristic of spontaneous speech). There were no direct methodological models to go by, since research into spoken language still was a relatively undeveloped area of research, even in an international perspective. The basic pattern was therefore set by devising a couple of manuals: *Manual för analys och beskrivning av makrosyntagmer* by Bengt Loman and Nils Jörgensen (Lund 1971), and *Manual för grammatisk beskrivning av talad och skriven svenska* by Ulf Teleman (Lund 1974). A shorter article was written by Bengt Loman concerning the prosodic analysis of the material, *Brott, paus, planering* (in *Skrifter utgivna av Nämnden för svensk språkvård* 48). The methods were devised so as to facilitate computer processing of large amounts of data. The manuals were naturally also used in composing the information for the magnetic tapes. The former manual contains rules for segmenting the speech stream into different kinds of syntactic units, macrosyntagms, including complete, incomplete and fragmentary sentences. This process of segmentation was the practical basis for the more detailed analysis to be undertaken. The detailed analysis that follows concentrates on the internal syntactic structures within these segments. However, during our work on the problem of the macrosyntagms we found that a detailed analysis of the macrosyntagms themselves proved extremely interesting. For instance, we were able to get information on the extent to which people speak – spontaneously – in complete, well-formed sentences, or in sentence fragments. And we could get data on the amounts of unintentional repetitions, false starts and incomplete units that occur in everyday speech. Also, the extent to which different social groups and speech in different situations varied in this respect proved to be an interesting area of research.

The prosodic analysis provides opportunities for in-depth descriptions of the role of pauses in spontaneous speech. To what extent do we use pauses within a sentence between clauses and between parts of speech? Where do we most typically pause? What can we learn about the process of speech planning from the position and frequency of pauses?

The aspects of language that are described according to Loman's and Jörgensen's manuals are basically not concerned with the grammar of Swedish as such. Strictly speaking, they fall outside the domain of syntax proper, and are concerned with the structure and function of speech production itself. They are concerned with the manner in which people manage to fulfil their linguistic intentions in spontaneous speech.

Teleman's manual, on the other hand, focuses directly on analyzing speech into grammatical categories. On the surface the method of analysis is reminiscent of traditional sentence parsing. The manual is not based on any particular grammatical theory, but with some additions to the computer programming it could relatively easily be adapted to the principles of generative grammar. Every word in every clause is analyzed with respect to its lexical characteristics and its syntactic function. All such data are kept on special records, which are then punched for computer processing and fed onto the magnetic tapes. From the information on the magnetic tape it is then possible to extract many different kinds of grammatical information, for instance, the functions of the attributive genitive, the different types of combinations of parts of speech within a clause, the use of the passive constructions, and the position of different kinds of adverbials in a clause.

An important aspect of our work has been to develop methods for data collecting and for classifying the material according to non-linguistic categories. Research into spoken language has to battle with an unfortunate paradox: the aim is to investigate how people speak spontaneously; this they do at their best when they feel that they are not being observed; on the other hand, there is no way to observe people's behaviour without observing it. Nevertheless, the general opinion among researchers is that the best source of material for investigations into spontaneous spoken language is the interview (on a familiar and interesting topic). Most of the material in *Talsyntax* therefore consists of interviews. The interviews were devised to look like sociological polls, and even used as such. This made our linguistic goal less perspicuous

to the people interviewed. All in all some 200 speakers of Swedish were interviewed in the towns of Borås, Trelleborg and Helsinki and in the district of Tornedalen. The choice of informants was made by chance, though we did take into account their socio-economic status, and tried to get an even representation of subjects representing different categories; in particular, we made use of a rough classification into professionals, businessmen, clerks/salesmen, and workers. The continuing research that has been done with our corpus has in fact focused on linguistic differences that might be socially governed. Also, we have attempted to make an even more delicate classification of the interviewed people, and a specially formatted social index has been made up for each person.

Another type of variation in spoken language is caused by the difference between more formal and more informal speech. In planning the recording sessions we could naturally make use of the generally accepted fact that people talk more formally the more conscious they are of somebody observing them, and the more they themselves are conscious of the importance of the linguistic form. A taping session was arranged where groups consisting mainly of professionals were gathered to partake in a debate on a departmental problem. The debate (with a chairman) was preceded by a warming-up session, where the participants were able to talk through the issues in the way of preparation. However, this part of the session was also taped, without the participants' knowledge.

As a whole the material contains examples of spoken Swedish in some typical speech situations: interviews, conversations, and debates. In addition to this the material also includes a number of tape recordings of authentic committee meetings both on a national and a more local level of administration. Here, too, the styles of the speakers were found to vary according to the roles they played at different points in the meetings: giving a speech, partaking in the general discussion, acting as chairmen, etc.

At a later stage we were able to broaden our investigations of the stylistic and functional variations of spoken language, in that texts and data from an investigation into the language of TV and radio were made available to us. This investigation has been started as a special project at the Department of Nordic languages of the University of Lund.

3. Empirical Results

The empirical results of the project have thus far been published mainly in a special series (Series C) within *Lundastudier i nordisk språkvetenskap* (Studentlitteratur, Lund). In particular, four Ph.D. theses should be mentioned:

Mirja Saari, *Talsvenska* (Helsinki 1975);

Nils Jörgensen, *Meningsbyggnaden i talad svenska* (Lund 1976) and *Underordnade satser och fraser i talad svenska* (Lund 1978);

Jan Einarsson, *Talad och skriven svenska* (Lund 1978); and

Godelieve Laureys, *De adverbiale bepaling in het Zweeds* (Gent 1979).

Results from work within the project have also been published in two volumes with the title *Språk och samhälle* (Gleerups, Lund 1972, 1974). A third volume with the subtitle *Social variation i talspråkets syntax* was published in 1976 and may be regarded as a final report from the project. Furthermore, the anthology *Barnspråk i klassamhälle* (Liber Läromedel, Lund 1974), edited by Bengt Loman, deals with questions of linguistic handicap and class-dirrerences in connection with the theories of the British sociolinguist Basil Bernstein. An attempt to apply the findings of the project in education is sketched in Loman's article *Miljö och språkförmåga* (in *Svensklärarföreningens årsskrift* 1973).

A theoretical starting-point for our work has been Bernstein's argument that there exist two codes which govern the use of language in the British middle class and working class, in the direction of an elaborate code, and a restricted code, respectively. Bernstein's theories have been of major importance in many countries, especielly in their pedagogical applications. The work within *Talsyntax* has partly been directed toward an investigation of the extent to which it is possible also in modern Sweden to find similar sociolinguistic oppositions, in syntax in particular. This we considered an urgent task, since the international debate around Bernstein's theories has suffered from a certain lopsidedness. The theoretical research has been far ahead of empirical verification, and has even at times been confused with empirically proven facts. The empirical investigations have been very few, and often restricted as to planning and purpose; for instance, there has been a lack of data on adult syntax.

It is a task both difficult and ungrateful to summarize the results of *Talsyntax* on this issue within the limited space I have at my disposal here. Particular data can be analyzed and interpreted in different ways. Still, the material does give an impression that the spoken language is surprisingly homogeneous in the areas we investigated. Not only is it true that different informants use basically the same stock of grammatical constructions; even the frequency figures for different phenomena are similar from one speaker to the next. A bird's-eye view of the material makes one want to talk about "a common grammar". Another general observation is that spoken language is much more rule-governed and grammatical than what one perhaps tends to think intuitively. There is a constant effort on the part of the speaker to express himself in complete and well-formed sequences. Mirja Saari's investigations have shown that about 70% of all the sentences in spontaneous speech are "correctly constructed"; even sentences that are "wrongly constructed" can decrease considerably in number if they are filtered through a set of editing rules, which allow you to overlook sequences that the speaker himself marks as having been wrongly constructed, for example by changes in his pitch range and speech tempo.

There are also many detailed differences between speakers of different groups. However, one has to be careful in making judgments about these data, especially in relation to linguistic class differences. The material is, after all, relatively limited, and the analyses were mainly focused on strictly linguistic issues. Further, the analysis has only been concerned with isolated sentences and their internal form. Neither the content or sense of the text, nor the function or communicative effect of the utterances has been considered. The way we conducted the tape-recordings has not allowed us to consider those extra-linguistic variables that always accompany speech: gestures, mimics, speech tempo, voice intensity for instance. Also, only to a very limited extent has it been possible to distinguish between ready-made, formulaic phrases, and on-the-spot-created sequences in spontaneous speech. Above all, there are no workable models for detailed analyses of the relation between language and the socio-economic structure of society.

Therefore the *Talsyntax* project should be seen primarily as a study of the general characteristics and peculiarities of spoken Swedish. As such our work will be of practical significance for the teaching of Swedish as a foreign language, and for perceptual judgments of different varieties of written and read-aloud Swedish.